Descriptive Analysis Insights

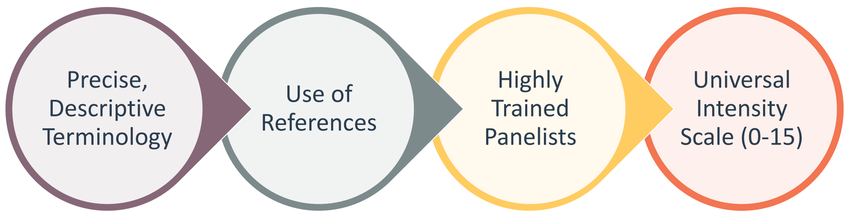
Pros and cons of using consensus vs. individual data collection with and without replication

Osdoba, K. E., Gibbons, S., & Retiveau Krogmann, A.
Sensory Spectrum, Inc. New Providence, NJ. USA. www.sensoryspectrum.com

Spectrum Descriptive Analysis

In today's food industry, efficiency of research methods is highly valued, though without sacrificing validity and reliability of data.

The Spectrum Descriptive Analysis is a recognized valid and reliable method to objectively document the sensory characteristics of products. Hallmarks of the method include:



In operation, the Spectrum Descriptive Analysis method is very flexible, and offers many variants, for example in type of data collection:

Individual	Consensus
• Panelists evaluate samples individually • No discussion • Profiles summarized as means • Sample comparison via statistical analysis	• Panelists evaluate samples individually • Group discussion with panel leader • Single profile recorded • Sample comparison via expert interpretation

Another area that offers flexibility is in the number of replications. Many descriptive methods encourage the use of replications. Spectrum Descriptive Analysis is not always as definitive in its recommendations.

The Questions

With many variants of the method available, many questions arise that this research attempts to answer:

Individual vs. Consensus – Is one Method Better than the Other?

What are the key similarities & differences between data obtained via consensus and data obtained individually?

Is the interpretation of the results similar with both methods?

When might it be appropriate to use one method over the other?

In what situation does individual data collection provide an advantage over consensus evaluation, or vice versa?

Replications – Are They Really Necessary?

What is the effect of replication on interpretation of descriptive results?

When is it appropriate to use 1 rep vs. 2 (or more) reps of descriptive evaluation?

Under what conditions does the gain in sensitivity provided by multiple reps outweigh the time investment that it requires?

Does it depend on which data collection method is used?

Methodology

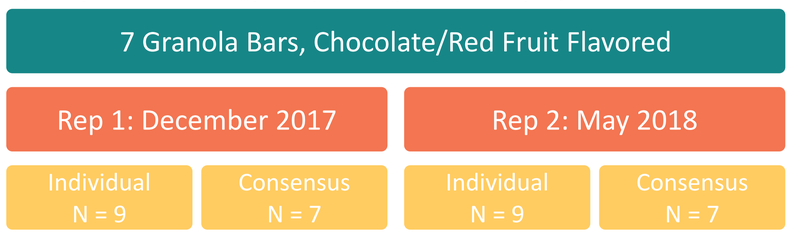
Product Category

Granola bars were chosen as an example of a complex product comprised of multiple components in varying proportions (i.e. grains, nuts, fruit pieces). Many of the granola bars in the sample set have high within-bar variability (i.e. different bites may have a different ratio of components) and high batch-to-batch variability. This category was chosen for this project to best illustrate the capabilities of both data collection methods for capturing these two types of variability.

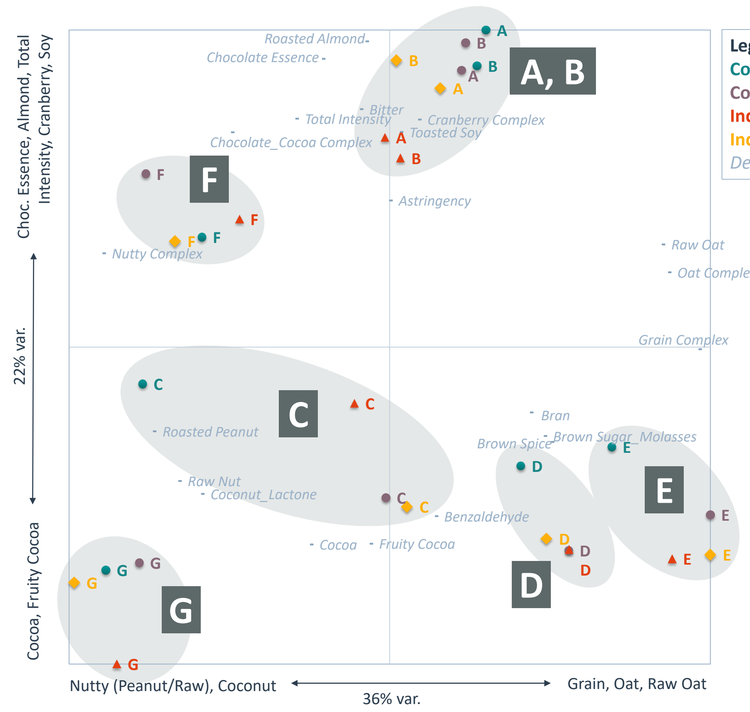
Panel Evaluations

Prior to evaluation, an orientation session was conducted to ensure panel alignment in techniques and terminology. All orientation and evaluation sessions were facilitated by a trained Spectrum panel leader. Panelists were presented with two unwrapped bars for each evaluation and instructed to take a minimum of three bites from both bars prior to forming an opinion and recording their rating for each balloted attribute.

Data were analyzed using methods appropriate to each objective including descriptive statistics, analysis of variance, principal component analysis, and correlations.



Key Findings – Individual vs. Consensus



At the Profile Level

Both individual and consensus data collection methods characterize samples in similar ways – each sample's four evaluations group together on the map at left.

Very good agreement in sample characterization is seen between methods and between reps. The same conclusions may be drawn from the sample set, regardless of data collection method or number of reps.

The results illustrate that both consensus and individual data are well able to capture the variability of a complex, multicomponent product like granola bars, with or without replication.

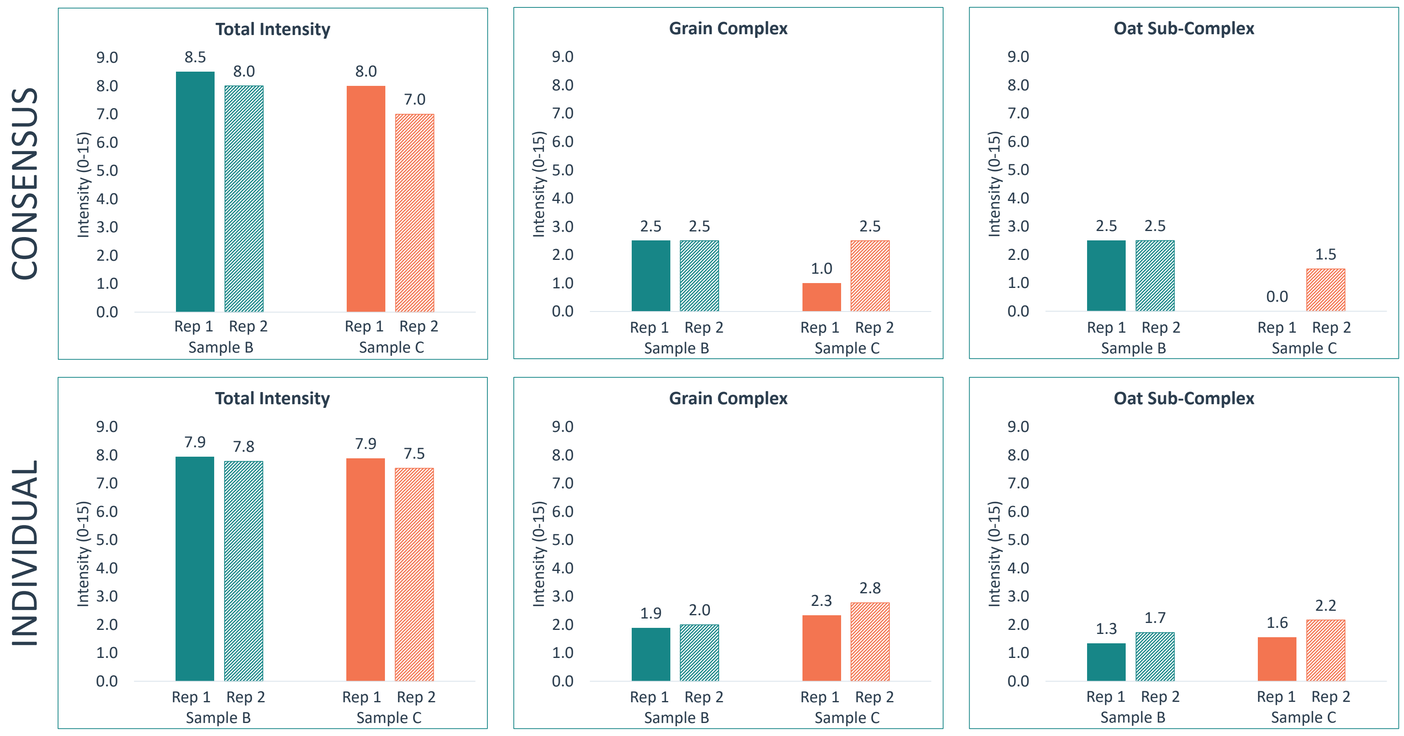
Sample C is the only sample for which wide variation is seen. The sample evaluated in the first rep was seen as higher in toasted soy and chocolate essence and lower in oat and fruity cocoa compared to the second rep, indicating higher batch-to-batch variability compared to the other samples in the set.

At the Attribute Level

High agreement & alignment between the two data collection methods are shown for 90% of attributes. Of the 38 attributes rated above threshold for at least one sample, high correlation is seen between consensus and individual data for 23 attributes. Considering 25 attributes for which both methods differentiate the sample set, high correlation is seen between consensus and individual data for 20 attributes. For seven attributes, only one method differentiated the sample set.

	Level of Alignment between Consensus and Individual *Number of Attributes with:*			**Total Number of Attributes (rated above threshold)**
Differentiation by:	High Alignment (r > 0.7)	Moderate Alignment (r > 0.5)	Low/No Alignment (r < 0.5)	
Both methods	20 √ √	3 √	*2 X*	**25**
One method	2 √	3 √	*2 X*	**7**
Neither method	1 √ √	1 √ √	4 √ √	**6**
Total	23	7	8	**38**

Key Findings – Effect of Replication



The use of replication is a recognized standard practice for many descriptive methods. Findings from this study reveal that replications may not always be necessary.

- Sample B's profile is stable across replications. Results are shown for three attributes varying in level of nuance. Results are aligned between the two data collection methods, suggesting that replications may not be necessary. Note that replications were performed months apart: Spectrum DA with a validated, high-performing panel is stable over time for consistent products.
- Sample C, however, shows more variability across replications: differences between replications are directionally aligned between the consensus and individual data collection methods, suggesting greater batch to batch variability for that sample.

Conclusions

Individual vs. Consensus – Is one Method Better than the Other? *It Depends!*

Using the Spectrum Descriptive Analysis method with a validated panel offers reliable, reproducible, and nuanced profiles, regardless of whether consensus or individual data collection is used.

Conclusions from both methods generally align, and demonstrate similar level of sensitivity to differences.

Some slight differences exist however:

- Consensus relies on decision process which maximizes the chance of documenting extremely small differences among samples.
- Individual data relies on individual decision process for allocating perception to attributes, creating slightly more noise in the data. Differences, when shown, are more likely to be meaningful.

Replications – Are They Really Necessary? *Sometimes, Not Always!*

Replications (for either data collection method) are useful for measuring and accounting for within-sample and/or batch-to-batch variability. Results suggest they are not always necessary, especially when samples are believed to have strong quality control and consistently deliver a similar sensory experience.

Recommendations	Use Consensus Method When…	Use individual Data Method When…
Panelist availability	Smaller pool of panelists (n=6-10)	Larger pool of panelists (n=8-12)
Logistical and Analytical Support	No specialized software for data collection / analysis needed.	Specialized software for data collection / analysis needed.
Panel Familiarity with the Category	Panel does not evaluate the category frequently (weekly). Discussion allows better clarification of terms in session.	Panel evaluates the category frequently (weekly). No discussion needed to clarify attribute allocation.
Panel leader skills	Panel leader is experienced and highly skilled at facilitating consensus discussion (group dynamics).	Panel leader is less familiar with facilitation techniques and group dynamics.
Study type	Exploratory in nature, such as category review or large variability study.	Studies intended to guard against falsely concluding for differences (e.g., shelf-life, ingredient or process change…) Claim substantiation – per ASTM guidelines.
Product Category	Categories where tiny differences / nuances may have large impact (e.g., coffee).	Categories where large individual differences in perception are suspected (e.g., non-nutritive sweeteners).

Do not replicate when…	Replicate a subset of the samples when…	Replicate all samples when…
You suspect within-sample variability to be relatively small. You expect differences among samples that are larger than within-sample variability.	You want to check the panel performance (use reps to highlight if the panel is consistent) You suspect some samples to be more variable and want to understand that variability. No specialized software for data collection / analysis needed.	Understanding sample variability is critical for all samples. You will be using the information to substantiate a claim – per ASTM guidelines.

sensoryspectrum